Who tweets about cancer? An analysis of cancer-related tweets in the USA

Dhiraj Murthy[1] and Macgill Eldredge[2]

Abstract

Cancer patients, family members and friends are increasingly using social media. Some oncologists and oncology centres are engaging with social media, and advocacy groups are using it to disseminate information and coordinate fundraising efforts. However, the question of whether such social media activity corresponds to areas with higher incidence of cancer or higher access to cancer centres remains understudied. To address this gap, our study compared US government data with 90,986 cancer-related tweets with the keywords 'chemo', 'lymphoma', 'mammogram', 'melanoma', and 'cancer survivor'. We found that the frequency of cancer-related tweets is not associated with mammogram testing and cancer incidence rates, but that the concentration of doctors and cancer centres is associated with cancer-related tweet frequency. Ultimately, we found that Twitter has value to cancer patients, survivors and their families, but that cancer-related social media resources may not be targeting locations that could see the most value and benefit. Therefore, there are real opportunities to better align cancer-related engagement on Twitter and other social media.

Keywords

Twitter, social media, cancer, telemedicine, healthcare access, melanoma, lymphoma, chemotherapy

Submission date: 1 October 2015; Acceptance date: 26 May 2016

Introduction

Modern social media such as blogs, Instagram, Facebook and Twitter have increased the ability of patients to find others with their conditions and to discuss treatment options, suggest lifestyle changes, and to offer support. Indeed, social network sites completely built around patients such as patientslikeme.com have been developed to link patients with a particular illnesses together and to provide researchers with rich patient data. The 140-character microblogging platform, Twitter, has been actively used by patients because of the medium's ubiquity and ability to connect with health professionals and well-informed patients. Twitter has a disproportionately high uptake amongst non-white Americans and Twitter is biased towards people aged 18–49 years.[1] In addition, Twitter use for American adults aged over 65 years doubled from 2013–2014.[1] These user demographics present new opportunities to engage with older Americans as well as with younger populations who may feel less anxiety using computer-mediated communication.[2] In some cases, Twitter and other social media have also enabled

traditionally isolated patients such as the physically disabled to connect with strangers with similar conditions.[3]

These new uses of social media have also led to more public disclosure of health information. For example, Lisa Bonchek Adams, a breast cancer patient, tweeted over 176,000 times with many of those tweets about her own cancer experience. Adams' cancer eventually spread to her bones and the mother of three from Connecticut died in April 2015. Her tweets intimately chronicled her cancer experience (e.g. on 4 January 2014: 'Very rough day here. Dizziness, weakness, pain. Need the tumours to shrink for relief. That will take time: chemo and radiation'.). Though some have questioned the utility of high levels of disclosure by

[1]Goldsmiths College, University of London, UK
[2]TripAdvisor, USA

Corresponding author:
Dhiraj Murthy, Department of Sociology, Goldsmiths College, University of London, New Cross, London, SE14 6NW, UK.
Email: d.murthy@gold.ac.uk; Twitter: @dhirajmurthy

cancer patients,[4] there are many cancer patients, survivors and their family members who use Twitter to find out about treatments and side effects, or as a place of support.[5] These emergent uses indicate that our privacy norms are shifting as a result, making studies of Twitter and cancer particularly timely.

The purpose of this article is to establish the value of Twitter to cancer stakeholders and then study the correlation between cancer incidence, quality of cancer centres and provision of doctors to determine whether cancer-related social media resources are being provisioned in the right locations. To do this, we collected six months of cancer-related tweets and investigated whether their frequency is tied to cancer incidence, mammogram testing rates and other cancer-related statistical data. Prior to studying our collected data as a whole, we randomly sampled tweets by 1000 unique users to discern user demographics and message content. The purpose of this approach is to offer initial thoughts on the value of Twitter to cancer stakeholders. Though we hypothesised that US states with higher cancer incidence would have more cancer-related tweets, what we found was that the concentration of doctors and cancer centres is more associated with a state's volume of cancer-related tweets. This article uniquely contributes to the cancer literature by providing evidence that Twitter has value to cancer stakeholders, but that value is not necessarily being provided in the 'right' locations.

Twitter and disclosure

As the Lisa Bonchek Adams case highlights, tweets by cancer patients can regularly involve high levels of self-disclosure. Cancer patients have been found to tweet about their cancer diagnoses in the moment as well as updating their network about how their chemotherapy has been going.[6] This creates different sociological expectations of health-related behaviour on social media. Specifically, these users may see Twitter as a diary or may be actively soliciting an audience response. Though tweets can be directed to individual users, many tweets are undirected, broadcast to a users' following and Twitter as a whole. This can place primacy on communiqués that are generally addressed. And because the length of tweets is restricted to 140 characters and mobile application (app) adoption is widespread, it is generally easier to tweet about cancer compared to blogging, another medium popular with cancer patients.[7]

The structure of Twitter blurs what we perceive as private/semi-public spaces (i.e. followers) into public spaces (depending on the reach and number of one's followers as well as the keywords in one's tweets). This change in social communication has implications for patients in that Twitter may be exponentially increasing shifts in social communication from the private to the public.[8] Twitter also follows the trend set by Facebook of 'frictionless sharing',[9] where many types of content are shared with little 'friction' on one's profile. Location information, a purchase and what one is reading can be shared with little or no effort. This auto-posting augments the general trend towards making personal information less private. In the case of health, fitness trackers and smart scales, such as Fitbit and Jawbone, can auto-tweet one's fitness activity and weight. A smart bra designed in Greece tweets 'Don't forget to check your breasts women' every time the bra is unclasped in an effort to remind women of their breast self-examination.[10]

Twitter and health

Twitter has had an impact on the ways in which health information and resources are shared. For example, live tweeting during surgical procedures has been used to enable family members to get full information during a procedure as well as to provide an educational tool for other surgeons and for students. In May 2013, Dr Dong Kim performed brain surgery to remove a tumour from a 21-year-old female at the Texas Medical Center while his staff tweeted and uploaded videos live during the operation.[11] Members of the public were tweeting and following the procedure. Students thousands of miles away in New Jersey actively followed the Twitter feed and asked doctors about how the tumour grew, surgical complications and other questions. Forms of live tweeting are becoming more common despite a variety of legal and ethical issues.

As the live tweeting during the surgery in Texas highlights, Twitter provides a unique opportunity for more accurate health information to be disseminated to a diverse and wide audience. This could include others with brain tumours to interested members of the public or as an educational resource.[12] Additionally, Twitter may be useful in informing the public about health outbreaks. As McNab[13] argues, 'one fact sheet or an emergency message about an outbreak can be spread through Twitter faster than any influenza virus'. Lastly, Twitter changes the relationship between health institutions (including individual doctors) and the public in that government-issued health warnings and advice can now be more interactive, enabling conversations with the institution or person tweeting that information (e.g. replying to a local government's health department). In this way, Twitter, can potentially foster better health outcomes as the public may feel that they are making an informed decision. For instance, tweets could encourage individuals

to schedule a colonoscopy or mammogram after interacting with health institutions or patients who beat cancer that was discovered at an early stage.

Twitter and similar social media also present new opportunities for patient support networks. Keller[14] investigated tweets with the #hypothyroidism hashtag and found that the medium encouraged interactions between hypothyroidism patients and ultimately increased their agency. Hawn[15] finds that those who are chronically ill are successfully using Twitter to form support communities, which encourage healthy behaviours amongst patients (e.g. nudging diabetes patients to exercise and eat well). Vance et al.[16] argue that Twitter lends itself to a 'medical support group format' and offer the example of a Twitter user who uses her timeline as a network for mothers of children who have attention deficit disorder. These examples highlight the ability of health-related hashtags to form ad-hoc support communities.

The conditions that tend to have the most active networks are usually chronic or life changing. For example, cancer patients are highly active on Twitter and some users insert the phrase 'cancer survivor' into their user profiles.[5] And tweets during breast cancer awareness month were found to have raised 'general awareness and fundraising' for breast cancer.[17] Some cancer survivors are keen to help other survivors and use the medium to accomplish this. The case of cancer networks on Twitter presents a glimpse of not only how doctors and health institutions are interacting with individuals, but also how these networks have an international reach and, most of the time, involve strangers, rather than strengthening existing offline relationships. Though a small number of doctors are on Twitter,[18] they are usually interacting with 'far-flung' colleagues or members of the public[19] rather than with their patients. However, oncology professionals increasingly see the use of Twitter as an 'unprecedented opportunity' for 'high-priority' clinical trials for cancer.[20] Butcher[21] argues that Twitter is 'transforming the cancer care community' by connecting patients with oncology professionals. And given Twitter's generally younger user base, the transformation is likely disproportionately affecting younger people, a group for which discussing cancer may be difficult.[22]

Though health professionals and government institutions tweet about cancer, most cancer-related tweets are patient-generated. Chou et al. argue that this is 'seen as more democratic and patient controlled, enabling users to exchange health-related information that they need and therefore making information more patient/consumer-centered'.[23] However, tweeting about personal or family diseases and seeking advice usually necessitates elements of trust in the Twitter network. For instance, some health recommendations may be contradictory or made by users without the requisite professional training or experience. Because of the uncensored and collaborative nature of Twitter, users may not be getting the best advice. That being said, as Tsuya et al.[24] found, patients are turning to Twitter for cancer-related treatment, diagnosis, and symptom information. Their work importantly highlights that the medium is 'useful for cancer patients to exchange ordinary information' important to them, but is not necessarily seen as such by the healthcare community.

Research questions

The purpose of these research questions is to study the correlation between cancer incidence/healthcare resources and cancer-related Twitter activity in order to evaluate whether health-care-related social media resources are producing value in the appropriate locations (e.g. US states with high levels of cancer incidence).

Research Question 1: Are US states with larger populations more likely to tweet about cancer?
This is presented as a research question because recent work has shown that dense networks of grassroots cancer communities can highly influence cancer-related discourse on Twitter.[25] This is despite the fact that population is generally one of the most significant variables for Twitter use.

Research Question 2: Do US states with higher cancer-incident rates have a higher frequency of cancer-related tweets? This is presented as a research question because studies have shown that the incidence of other health-related behaviours such as influenza and alcohol consumption can be associated with tweet frequency.[26]

Research Question 3: Are US states with large populations of Twitter users more likely to tweet about cancer? Like Research Question 1, the purpose of this research question is to establish the influence of population versus local social media dynamics such as grassroots movements or particularly vocal Twitter users such as Bonchek Adams.

Research Question 4: Are US states with larger populations of young people more likely to tweet about cancer? This is presented as a research question because studies have shown that Twitter use is biased towards young people.[1] This research question aims to understand the role of age in cancer-related tweeting frequency.

Research Question 5: Does a greater concentration of doctors and cancer centres in a US state affect the frequency of cancer-related tweets? This is presented as a research question because there have been substantial efforts to increase Twitter activity within the
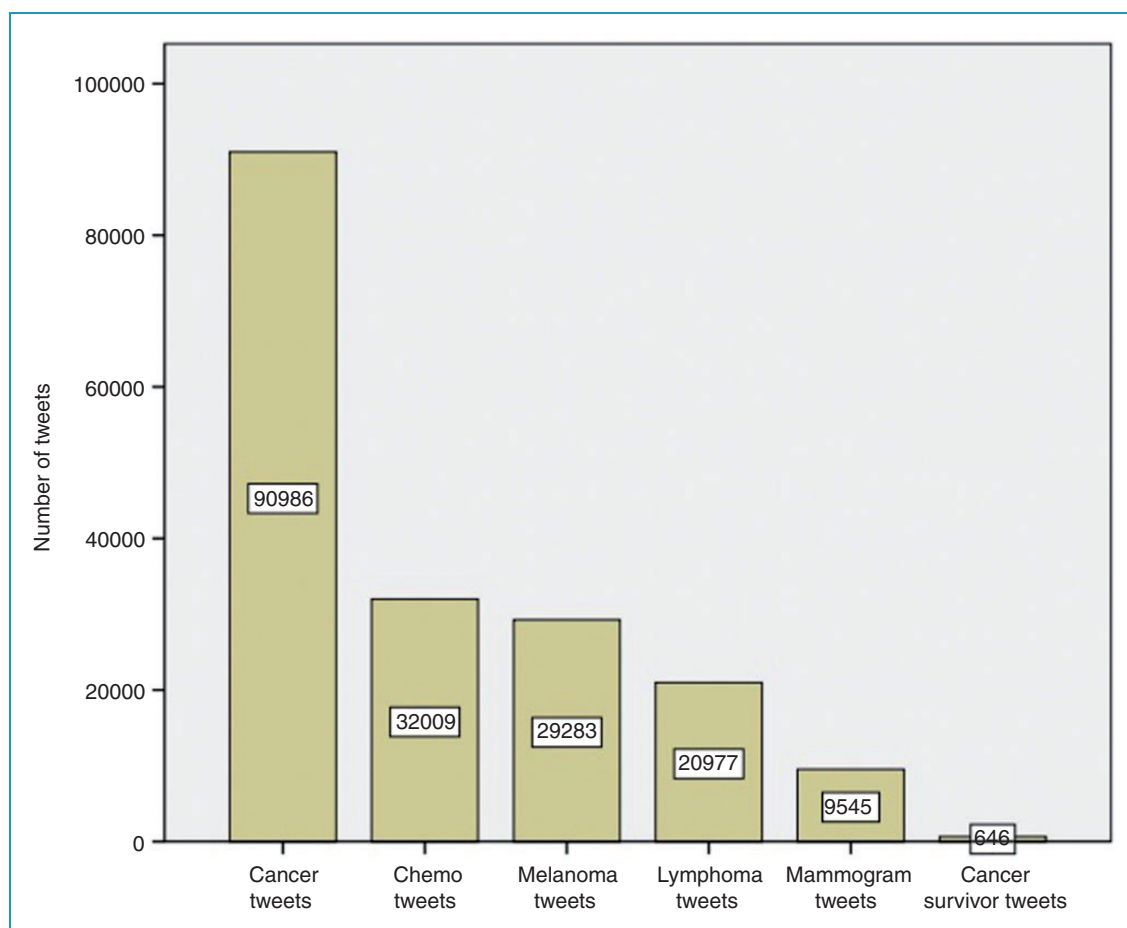
**Figure 1.** Frequency of tweets by cancer-related keywords.

oncology community, including guidance for the use of social media in oncology practice.[27] This research question is based on the premise that clusters of oncologists and oncology centres could be amplifying the dissemination of cancer-related messages amongst users in a US state.

Research Question 6: Does proximity to highly ranked cancer centres affect the frequency of cancer-related tweets? This is presented as a research question as top cancer centres such as the MD Anderson Center in Texas launched aggressive social media campaigns, including specific Twitter-based campaigns involving tweeting by top oncologists.[28]

Methods

A total of 90,986 tweets were collected from December 2010–May 2011 by directly querying the Twitter Streaming application programming interface (API) for the keywords: 'chemo', 'lymphoma', 'mammogram', 'melanoma', and 'cancer survivor'. Collection rates by keyword were approximately 65–82.2% of tweets that would have been returned with full firehose

API access. We had trial access to the full 'firehose' Twitter feed via discovertext.com for three days after our data collection period (15–17 October 2011). This enabled us to get a snapshot of what percentage of tweets, we would likely have captured during our data collection period. Four API queries were granted — 'chemo', 'lymphoma', 'melanoma' and 'mammogram'. We simultaneously restarted our data collectors during this period and compared our collection rates versus those with firehose. We found that we collected 65% of tweets for 'chemo', 81.5% for 'lymphoma', 82.2% for 'melanoma' and 68.2% for 'mammogram'. Frequencies for tweets collected by keyword are illustrated in Figure 1. All tweets were stored in a structured query language (SQL) database that included date, time, user location (including latitude/longitude coordinates if provided), tweet text and other JavaScript object notation (JSON)-returned attributes. Though significant numbers of tweets were collected for cancer-related keywords, location could not be verified for many (particularly in the case of 'cancer survivor'). Not all user locations have global positioning system (GPS) location coordinates attached to tweets (less
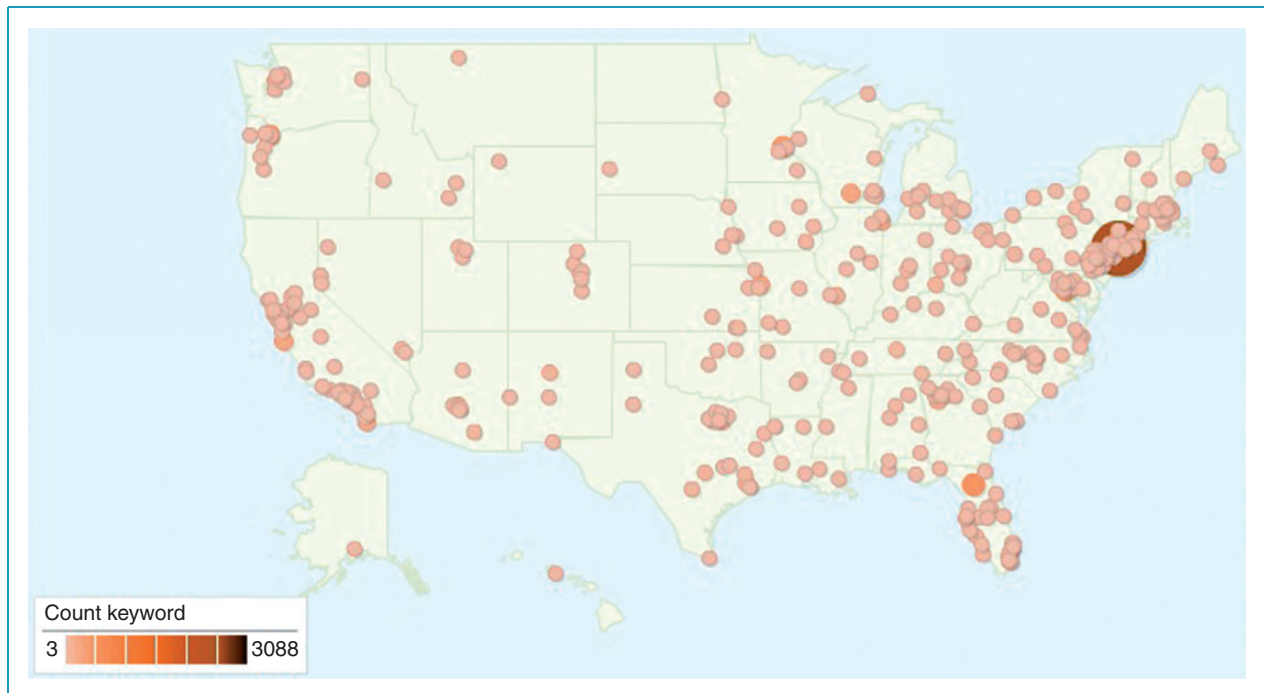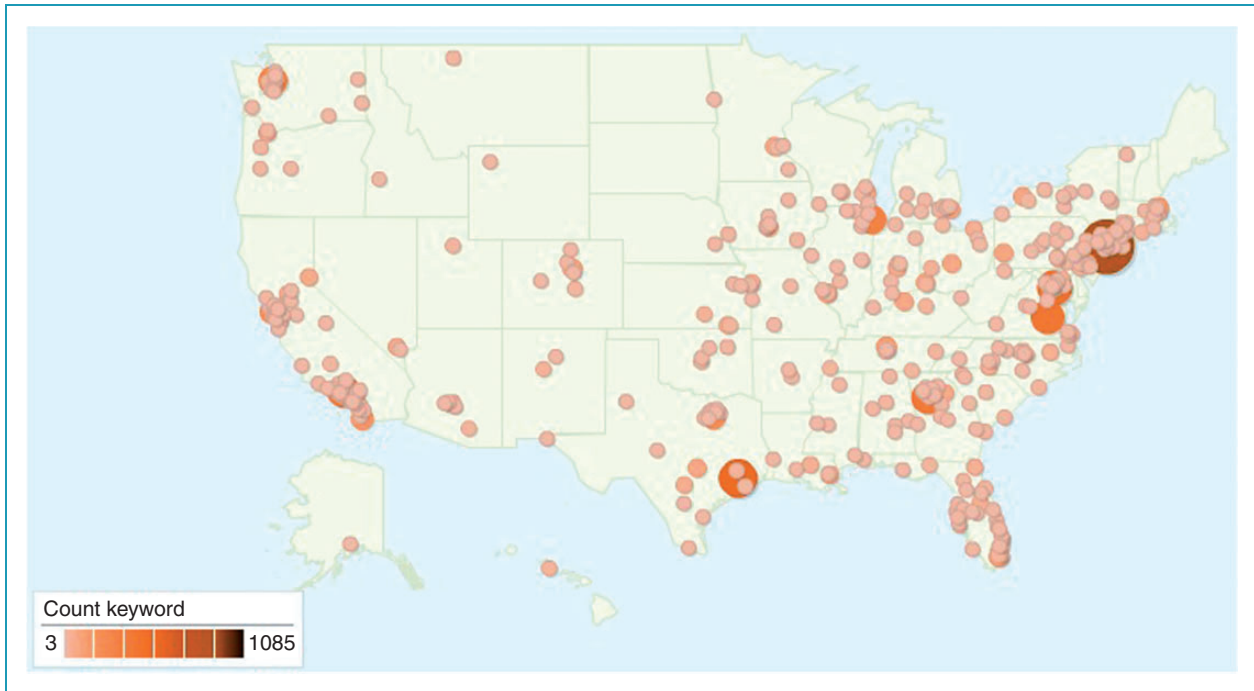
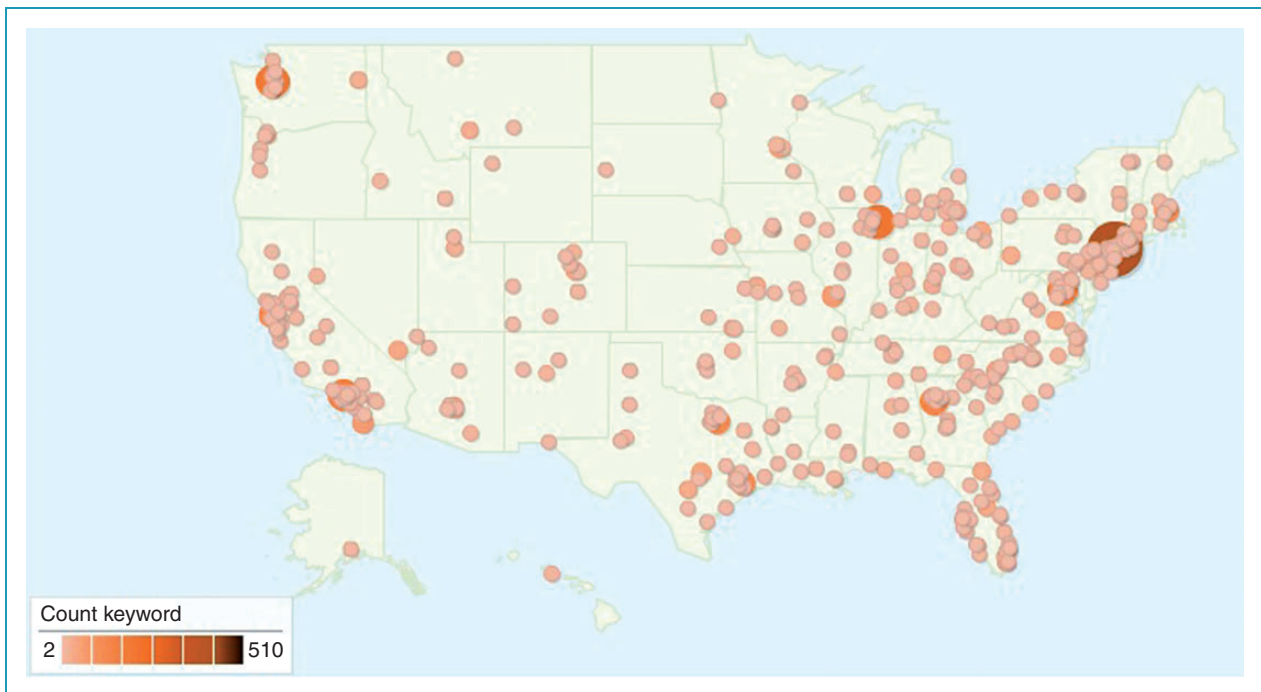**Figure 2.** Frequency of 'melanoma' tweets by USA location.

than 1% of our dataset had location). This is consistent with Graham et al.'s finding of 0.7% of tweets they sampled in 2011 to have exact GPS coordinates.[29] Like Graham et al., we implemented a location-inferring procedure to augment GPS data. Tweets without location coordinates, but with location information were 'cleaned'. If users specified a location in their profile, we passed this data to the Yahoo! PlaceFinder API. If a match was found, the returned location data was stored in a standardised 'city, state code, country code' format (e.g. Miami, FL, US). Raw location data not fitting the above procedure for coordinates were then passed in their entirety to the Yahoo! PlaceMaker API as it recognises zip/postal codes, country/state codes, and some colloquial names, such as 'The Windy City' and 'The Sunshine State'. These data were then used to augment our GPS data and allowed us to filter US cities in terms of cancer-related tweets (see Figures 2–4). We then bucketed tweets by US city and state to correlate with city- and state-level Twitter data and government indicators. Because location was authenticated in this robust way, the volume of tweets studied was a subset of the total tweets with location data collected. However, this method allowed us to derive some location information for 66.8% of tweets, with 20.6% of tweets categorised at the US-state level. This is similar to Burton et al.'s location reliability of 15.35–17.13%.[30]

We used US government health data[31] and US census population data[32] to evaluate whether top Twitter cities had a higher level of correlation to cancer-related tweets and whether the percentage of women over 40 who have had mammograms in the last two years is correlated with the volume of 'mammogram' tweets. Further correlations were conducted at a state level (for which government data exists for incidence of any cancer, melanoma, and lymphoma) as well as for the ranking of cancer centres, which was done by averaging the US News and World Report score for all cancer centres in each state. Twitter population was derived using data from a social media analytics tracking service, Hubspot.[33] We created the Twitter variables by population to correspond with cancer incidence rates (i.e. per 100,000) and these were done by the state level – again to be able to be compared with state-level government statistics. Given that cancer incidence rates are regularly correlated in the oncology literature using the Pearson's statistic[34] as well as previous work on Twitter and health,[35–37] we determined that Pearson's was most appropriate for our comparisons. Our data was organised by state and included variables for population, age, various cancer incidence rates, healthcare resources, quality of cancer care centres and Twitter-derived data amongst other things. All data were imported into SPSS, where both Pearson's and Spearman's rho bivariate correlations were performed. Pearson's and Spearman's values were found to be similar, providing an additional check. For example, using the data in Table 1, the Pearson values are: Population with Twitter population (Pearson: $r = 0.977$, $p < 0.01$ and Spearman's

Figure 3. Frequency of 'lymphoma' tweets by USA location.

Figure 4. Frequency of 'mammogram' tweets by USA location.

rho $= 0.958$, $p < 0.01$), Population with cancer tweets (Pearson: $r = 0.920$, $p < 0.01$ and Spearman's rho $= 0.954$, $p < 0.01$), and Twitter population with cancer tweets (Pearson: $r = 0.951$, $p < 0.01$ and Spearman's rho $= 0.949$, $p < 0.01$). Pearson values are the values reported in this study.

Before undertaking quantitative analysis of the nearly 100,000 tweets we collected, we randomly

Table 1. Correlations between US state of population,[32] Twitter user population by US state[38] and cancer-related tweets.

		Population	Twitter population	Cancer tweets
Population	Pearson correlation	1	0.977[a]	0.920[a]
	Sig. (two-tailed)		0.000	0.000
	n	51	50	51
Twitter population	Pearson correlation	0.977[a]	1	0.951[a]
	Sig. (two-tailed)	0.000		0.000
	n	50	50	50
Cancer tweets	Pearson correlation	0.920[a]	0.951[a]	1
	Sig. (two-tailed)	0.000	0.000	
	n	51	50	51

[a]Correlation is significant at the 0.01 level (two-tailed).

sampled 1000 tweets from unique users and went through these manually to identify the type of Twitter user (whether they were a cancer patient, survivor, family member, healthcare centre, etc.). As we were also interested in exploring the content of these messages, we manually classified these tweets by whether they related to news, health information, fundraising, etc. Table 2 illustrates the coding categories used. The codebook was developed by the first author using a grounded theory approach.[39] Specifically, tweet content and user types were qualitatively observed using a process of systematic note taking and an emergent codebook was developed. Coding of cancer-related content on Twitter has been successfully done previously[40] and this work used similar coding categories, including personal experience and advice.

The randomly sampled tweets were required to have unique users. Both tweets and users were coded by two research assistants according to the codebook categories listed in Table 2. Intercoder reliability was Cohen's kappa $\kappa = 0.92$ for user type and $\kappa = 0.88$ for message type. Because our grounded theory approach yielded extensive note taking for each emergent category, the codebook was able to reflect a clear rubric, which we believe supported this high level of intercoder reliability.

We were interested in determining whether there are differences in tweeting patterns by user group type (e.g. healthcare centre, cancer survivor, cancer patient, family member, journalist, etc.) For the keywords collected (specifically chemo, cancer survivor, lymphoma, melanoma and mammogram), we produced separate cross tabulations and calculated chi-square statistics in order to explore differences in tweeting behaviour by groups. As these are nominal (categorical) variables, the chi-square statistic is a robust method to explore group-level differences by user type and tweet keyword variables. In the case of the user type variable, values were assigned for each user type. For the tweet keyword variables, '0' indicated a lack of a presence of the keyword in the tweet and '1' indicated that the keyword was present in the tweet.

Results

Pilot study

The purpose of this pilot study was twofold: (a) to establish the use and value of Twitter in cancer and (b) to get preliminary demographic data on who is tweeting about cancer. For this pilot, we first randomly sampled 1000 tweets from unique users from a total of 90,986 collected tweets and, as discussed in the Methods section, these data were coded by hand. In terms of use and value of Twitter in the context of cancer, we found that tweets contained news-related content 23.3% of the time and 'personal' content (e.g. personal anecdotes, stories or jokes) were found in 13.4% of the tweets. Excluding cancer-related tweets by non-English speaking and spam/malicious user types, we found that the most common users tweeting about cancer were 'Family member or friend' (49.3%), 'Cancer patient' (8.7%), 'Cancer survivor' (8.7%), 'Pet cancer' (7.4%), 'News organisation/journalist' (10.1%),

Table 2. Pilot study codebook categories.

User type	Message type
1 = Healthcare centre (hospitals, clinics, etc.)	1 = News
2 = Family member or friend (patient/survivor)	2 = Clinical trials, drug releases, etc.
3 = Pet cancer (animal-related cancer user)	3 = Advice giving (treatment, drugs, etc.)
4 = Cancer patient	4 = Advice asking (treatment, drugs, etc.)
5 = Cancer survivor	5 = Support giving (e.g. 'Hang in there')
6 = News organisation/journalist	6 = Support asking ('Pray for me'; 'I need support')
7 = Medical researchers/institution	7 = Health information/health advocacy
8 = Doctor	8 = Personal (stories, jokes, anecdotes)
9 = Medical professional (non-doctor)	9 = Advertisements (non-fundraising)
10 = Celebrity	9 = Fundraising related
11 = Non-English speaking user	11 = Non-English language tweet
12 = Robot (aggregator, automated, not spam)	12 = Other
13 = Suspended/removed/missing/spam	
14 = Other	

'Healthcare centre' (6%), and 'Celebrity' (4.7%). The highest frequency who tweeted about personal stories and experiences were 'Family member or friend', 'Cancer patient' and 'Cancer survivor' users (68%, 84% and 61% respectively). Interestingly, 'Pet cancer', users who had pets with cancer or pets who are survivors were 7.4% of the sample, a figure not far off from cancer patients or survivors. News organisations and journalists tweeted about news 60% of the time while healthcare centres had a balance of both news and health information (44% and 33% respectively). Celebrity users tweeted about supporting cancer research and patients 43% of the time and specifically about fundraising 29% of the time. This exploration of user types and messages indicates that cancer-related tweets are about six times more likely to come from a family member or friend of a cancer patient than from a cancer patient or cancer survivor. In addition, journalists are tweeting more than cancer patients and survivors.

For chemo tweets, the groups with significant differences in tweeting behaviour were family members and cancer patients ($\chi^2 = 15.884$, $p \leq 0.005$ and $\chi^2 = 16.511$, $p \leq 0.005$ respectively). Though these groups are perhaps the most likely to tweet about personal stories and these stories, anecdotes and jokes often reflected their experiences with chemotherapy, it is noteworthy that cancer survivors were not found to have significant differences in tweeting with 'chemo' ($\chi^2 = 1.182$, $p > 0.1$). For lymphoma, the groups with significant differences in tweeting behaviour were pet cancer ($\chi^2 = 6.802$, $p \leq 0.01$). This reflects the fact that lymphoma is the 'most common life-threatening cancer in dogs'.[41] However, surprisingly, melanoma, which is also common in pets, was not significant for pet cancer users ($\chi^2 = 0.247$, $p > 0.1$). Unsurprisingly, cancer survivors exhibited differences in tweeting with the keyword 'cancer survivor' compared to other user groups ($\chi^2 = 14.981$, $p \leq 0.005$). Unexpectedly, the only other user type that also exhibited significant differences within the 'cancer survivor' keyword data were automated bot accounts ($\chi^2 = 10.068$, $p \leq 0.005$). Bots were more likely to tweet using this keyword than other user types, not including cancer survivors themselves. Bots also had significant difference in their tweeting of the 'chemo' keyword ($\chi^2 = 18.073$, $p \leq 0.005$). Our findings suggest that cancer-related tweets by bots are generally used to advertise (e.g. selling products to cancer survivors).

The pilot study results indicate that cancer patients, survivors and their family members felt comfortable sharing personal stories of their experience with cancer. These results also provide evidence that Twitter has value to cancer patients, survivors, family members,

Table 3. Correlations including mammogram[31] and Twitter city rank[33] data.

		Pct. women mammogram test 2 years	Twitter rank	Cancer tweets	Sum of mammogram
Percentage of women over 40 with mammogram in the last two years	Pearson correlation	1	−0.144	0.042	0.030
	Sig. (two-tailed)		0.353	0.458	0.602
	n	310	44	310	310
Twitter rank	Pearson correlation	−0.144	1	−0.493[a]	−0.592[a]
	Sig. (two-tailed)	0.353		0.000	0.000
	n	44	50	50	50
Cancer tweets	Pearson correlation	0.042	−0.493[a]	1	0.958
	Sig. (two-tailed)	0.458	0.000	316	0.000
	n	310	50		316
Sum of mammogram	Pearson correlation	0.030	−0.592[a]	0.958[a]	1
	Sig. (two-tailed)	0.602	0.000	0.000	
	n	310	50	316	316

[a]Correlation is significant at the 0.01 level (two-tailed).

healthcare providers and journalists. If the random sample indicated a majority of spam or journalists, for example, our research questions would have evolved differently as the value of Twitter to cancer stakeholders may have been minimal or unclear.

Location-filtered tweets

Large cities such as New York, Los Angeles, Washington, Chicago and Houston have the highest frequency of cancer-related tweets, but much smaller cities such as Gainesville, Florida were in the top 15 cities of highest frequency cancer-related tweets. Because our data includes the location of tweets (GPS data plus user-derived location data), we were able to compare the frequency of cancer-related tweets with mammogram test uptake and cancer incidence rates (including all cancers as well as melanoma and lymphoma specifically). We found that the frequency of cancer incidence in a state is not significantly correlated with cancer-related tweets. As Table 3 illustrates, the sum of cancer tweets (the aggregate set of all cancer-related tweets) and mammogram tweets are both correlated with the top US Twitter cities ($r = 0.493$; $p \leq 0.01$ for the sum of cancer tweets and $r = 0.592$, $p \leq 0.01$ for the sum of mammogram tweets). These findings support the hypothesis that location is more correlated with cancer-related tweets than government reported health data. In other words, cancer-related Twitter activity was not found to correspond with mammogram test and cancer incidence rates.

Empirical studies of Twitter data reveal that the use of Twitter in health-related contexts is tied to broader trends in Twitter usage rather than associated with the incidence of particular diseases or medical conditions in an area.[6] Specifically, we previously found that the location of Twitter users posting tweets containing the keyword mammogram were strongly correlated with the cities ranked highest in tweet frequency rather than cities which had the highest incidence of mammogram testing.[6] In the case of mammogram tweets, our data suggest that the actual topic of the messages did not significantly influence the distribution of tweets. Rather, we found that the location of the user was most important. However, studies looking at health epidemics and Twitter show a high correlation of tweets and specific epidemics including H1N1[42] and influenza.[43,44] What this reveals is not only Twitter's utility to alerting us of pandemics, but also its demographic bias. For pandemics, this is not highly relevant as relative frequency is analysed. However, when tweets are compared to government statistics on cancer incidence, these data do not show a strong correlation. In other words, pandemics such as influenza can cross

Table 4. Correlation table by US state for age,[32] cancer incident rate[46] and cancer-related tweets.

		% Under 18	% Over 65	% Over 25 with bachelor's degree	Cancer tweets per 100,000	Cancer incident rate (per 100,000 persons)
% Under 18	Pearson correlation	1	−0.632[a]	−0.350[b]	−0.373[a]	−0.570[a]
	Sig. (two-tailed)		0.000	0.012	0.007	0.000
	n	51	51	51	51	49
% Over 65	Pearson correlation	−0.632[a]	1	−0.211	−0.126	0.336[b]
	Sig. (two-tailed)	0.000		0.138	0.379	0.018
	n	51	51	51	51	49
% Over 25 with bachelor's degree	Pearson correlation	−0.350[b]	−0.211	1	0.630[a]	0.106
	Sig. (two-tailed)	0.012	0.138		0.000	0.468
	n	51	51	51	51	49
Cancer tweets per 100,000	Pearson correlation	−0.373[a]	−0.126	0.630[a]	1	0.002
	Sig. (two-tailed)	0.007	0.379	0.000		0.988
	n	51	51	51	51	49
Cancer incident rate (per 100,000 persons)	Pearson correlation	−0.570[a]	0.336[b]	0.106	0.002	1
	Sig. (two-tailed)	0.000	0.018	0.468	0.988	
	n	49	49	49	49	49

[a]Correlation is significant at the 0.01 level (two-tailed).
[b]Correlation is significant at the 0.05 level (two-tailed).

many demographic boundaries, where other diseases usually have significant variance based on demographic factors such as race and income.
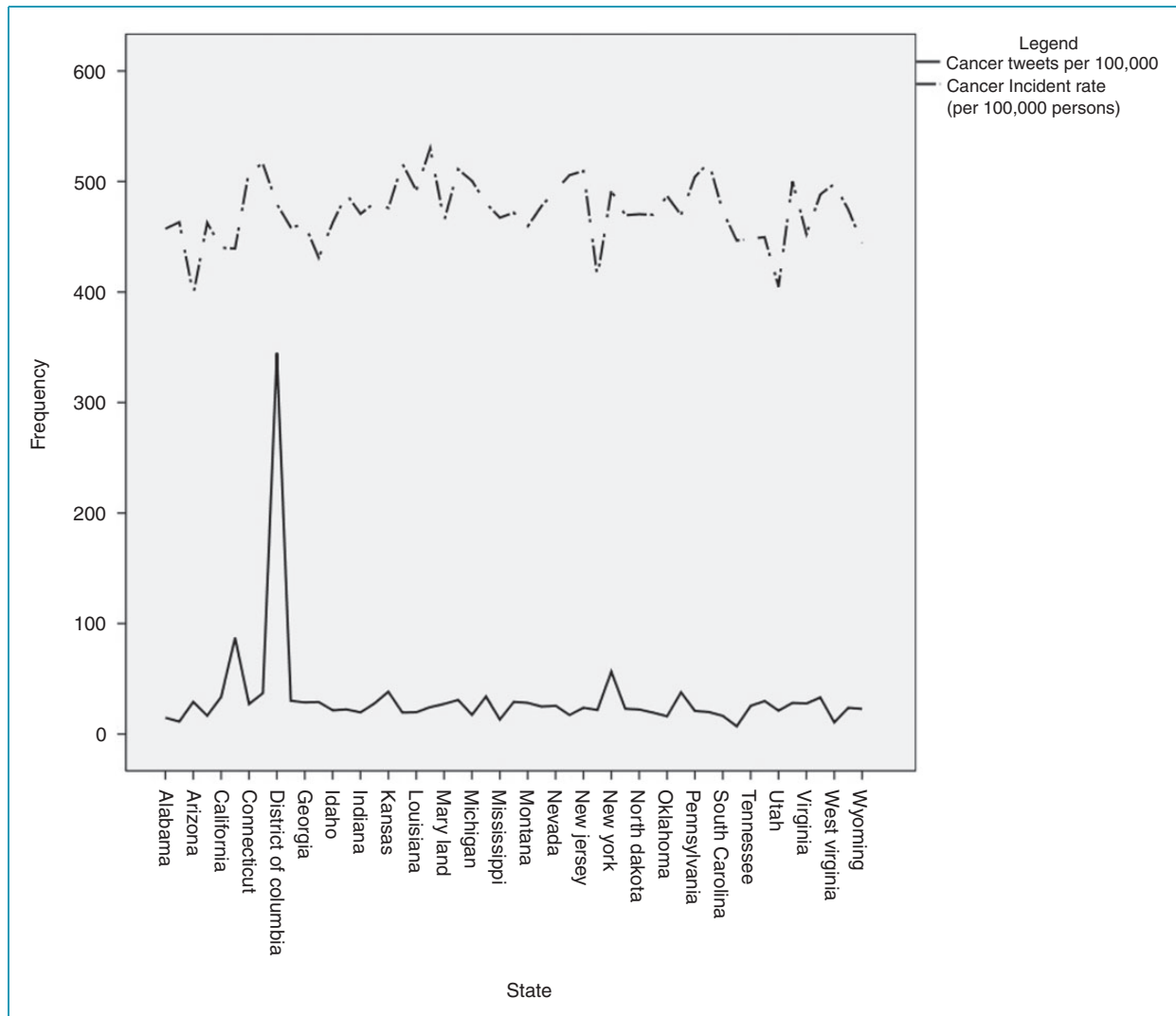
US cancer-related tweets

We examined the frequency of cancer-related tweets by American states and then explored correlates such as cancer incident rate. The frequencies of cancer-related tweets were sorted by keyword ('chemo', 'melanoma', 'mammogram', 'lymphoma', and 'cancer survivor') and their distribution is illustrated in Figure 1. Figures 2–4 illustrate the geographical distribution of cancer-related keywords by US state.

We found the following results regarding our research questions. Research Question 1 asks whether US states with larger populations are more likely to tweet about cancer. We found that the population of a state is highly correlated with the frequency of cancer-related tweets ($r = 0.920$, $p \leq 0.01$; see Table 1). This indicates that American states with larger populations are more likely to tweet about cancer. This supports the

literature, which indicates that tweet frequency and population size are associated.[45] Large populations generally have higher levels of Twitter use and this follows in the case of Twitter and cancer as well. Research Question 2 asks whether US states with higher cancer incident rates have a higher frequency of cancer-related tweets. Contrary to what we hypothesised, US state cancer incident rate is not correlated with cancer tweets ($r = 0.002$, $p > 0.05$; see Table 4). Indeed, as Figure 5 illustrates, the lines representing the incidence rate of cancer and cancer-related tweets show no discernible correlation at all (with the exception of the District of Columbia).

Research Question 3 asks whether US states with large populations of Twitter users are more likely to tweet about cancer. In Research Question 1, we found that states with larger populations were found to tweet more about cancer. Similarly, in Research Question 3, we found that US states with larger populations of Twitter users were likely to have a (relatively) large frequency of cancer-related tweets ($r = 0.951$, $p \leq 0.01$; see Table 1). Research Question 4 asks whether
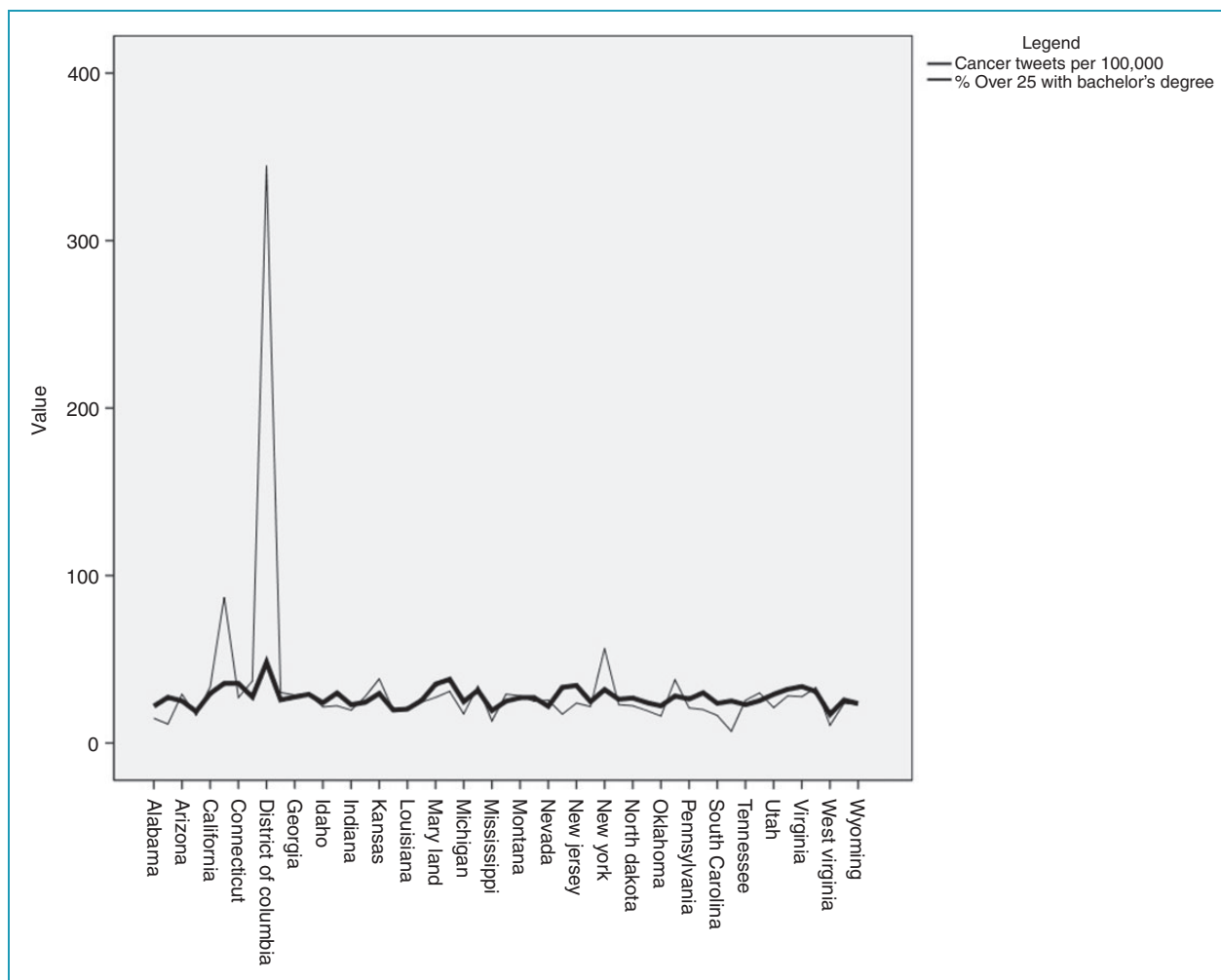
Figure 5. Frequency of cancer-related tweets and cancer incidence in US states.

US states with larger populations of young people are more likely to tweet about cancer. We found that cancer-related tweeting varied depending on age parameters. States with a larger percentage of people under the age of 18 years were negatively correlated with cancer-related tweets ($r = -0.373$, $p \leq 0.01$; see Table 4). However, states with a larger percentage of people over the age of 65 years saw no significant correlation with cancer-related tweets. The variable relevant to this research question with the highest correlation to cancer-related tweets was percentage of individuals over 25 with bachelor's degrees ($r = 0.630$, $p \leq 0.01$; see Table 4). In other words, states with educated individuals aged over 25 years are most likely to be responsible for posting cancer-related tweets. Figure 6 illustrates a strong correlation through a line chart, which shows that cancer-related tweets and percentage of individuals over 25 with bachelor's degrees have very similar curve patterns, indicating that cancer-related tweets are more frequent in highly educated, adult populations.

Tweet volume across all cancer-related keywords increased over time (see Figure 7) and would most likely be at much higher levels today, given the rapid growth of Twitter.[47] Rather than being attributable to a mutually exclusive interest in using Twitter for health-related social communication, the rate of increase is explained by an increase in the growth of Twitter over the period and a more general trend of exponential user uptake of Twitter.[48]

Both the variables for percentage of population under the age of 18 years and over the age of 65 years have negative correlations with the number of cancer tweets (note: the percentage of population over 65 years is not statistically significant, but is negative). Additionally, we found that the number of cancer

Figure 6. Frequency of cancer-related tweets and percentage of state population over 25 years old with a bachelor's degree.

incidents per 100,000 people is positively correlated with the percentage of population over 65 years of age ($r = 0.336$, $p \leq 0.05$; see Table 4). This suggests that despite the increased rate of cancer incidence in older populations, these populations are actually tweeting less than other populations. This further supports the finding that cancer-related tweets are not correlated with cancer incidence, but rather they are simply a result of the overall Twitter populations. In other words, the reason cancer-related tweets are negatively correlated with populations under 18 years of age and over 65 years of age is that the population of all Twitter users reflects this trend. It should be noted that at the time of our data collection, 87% of the Twitter population was comprised of users between the ages of 18 and 55 years.[49]

Research Question 5 asks whether a greater concentration of doctors and cancer centres in a US state affect the frequency of cancer-related tweets. We found that the concentration of doctors in a location reveals a strong correlation with the frequency of

cancer tweets ($r = 0.787$, $p \leq 0.01$; see Table 5) as well as the frequency of cancer centres ($r = 0.526$, $p \leq 0.01$; see Table 5). However, because the number of cancer centres is also correlated with the number of doctors, this suggests that there is most likely collinearity between the two variables and that they are both indicating the same correlation. However, between these two variables, we can conclude that states with a large number of doctors per capita – and probably more specifically cancer specialists – have a tendency to tweet more using specific cancer keywords.

Research Question 6 asks if proximity to highly ranked cancer centres affect the frequency of cancer-related tweets. We found that the average US News and World Report score for all cancer centres in each state does not have a significant correlation with the number of cancer-related tweets per 100,000 ($r = 0.267$, $p > 0.05$; see Table 6). This suggests that the quality of the cancer centres in a state does not influence the volume of cancer-related tweets as much as the actual number of centres (see Table 5). Of course,
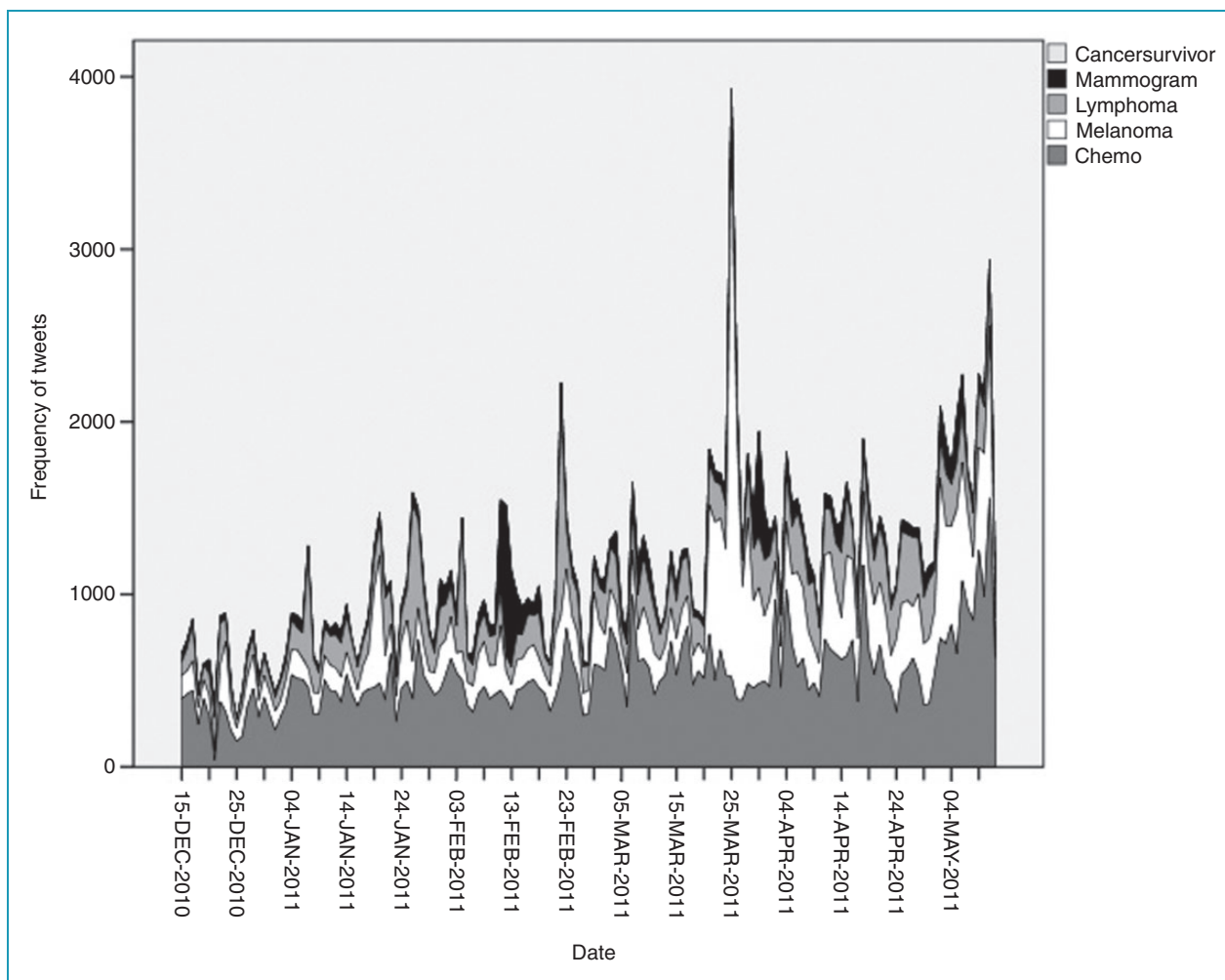
Figure 7. Frequency of cancer-related tweets over time.

Table 5. Correlations between concentration of doctors and cancer centres correlated with cancer tweets.

		Doctors per 100,000	Cancer tweets per 100,000	Cancer centres per 100,000
Doctors/100,000 residents	Pearson correlation	1	0.787[a]	0.526[a]
	Sig. (two-tailed)		0.000	0.000
	n	51	51	51
Cancer tweets per 100,000	Pearson correlation	0.787[a]	1	0.452[a]
	Sig. (two-tailed)	0.000		0.001
	n	51	51	51
Cancer centres per 100,000	Pearson correlation	0.526[a]	0.452[a]	1
	Sig. (two-tailed)	0.000	0.001	
	n	51	51	51

[a]Correlation is significant at the 0.01 level (two-tailed).

Table 6. Correlations between concentration of doctors, ranked quality of cancer centres and cancer tweets.

		Doctors per 100,000	Cancer tweets per 100,000	Average cancer centre score
Doctors/100,000 residents	Pearson correlation	1	0.787[a]	0.443[a]
	Sig. (two-tailed)		0.000	0.001
	n	51	51	51
Cancer tweets per 100,000	Pearson correlation	0.787[a]	1	0.267
	Sig. (two-tailed)	0.000		0.058
	n	51	51	51
Average cancer centre score	Pearson correlation	0.443[a]	0.267	1
	Sig. (two-tailed)	0.001	0.058	
	n	51	51	51

[a]Correlation is significant at the 0.01 level (two-tailed).

cancer-related tweeting trends are not independent of general tweeting trends and broader Twitter networks.[50] For example, states with larger populations between 18-65 years have historically shown an inclination for tweeting more than other states and are more likely to tweet using the cancer keywords as well. States with higher cancer rates showed no correlation with higher cancer-related tweet rates, and the only significant correlation found was between the frequency of doctors or the frequency of cancer centres and the frequency of cancer related tweets (see Table 5). However, our results indicate that states that could benefit from social media engagement due to higher incidence rates are perhaps not the locations where healthcare-related social media resources are being focused.

Conclusion

The purpose of this study was to demonstrate the value of Twitter to cancer patients, survivors, family members and health care professionals/institutions and to then ascertain whether social media is engaging locations with high cancer incidence rates. By utilising a range of research questions, we were able to study population effects (in terms of both general and Twitter populations) as well as the significance of highly ranked cancer centres to cancer-related Twitter activity. Contrary to what we hypothesised, cancer-related tweets are not associated with the incidence of particular cancers in an area. Specifically, we found that the location of Twitter users posting cancer-related tweets was significantly associated with (a) the most populous American states and (b) those states with

larger numbers of Twitter users. Cancer-related tweets were not significantly correlated with states that had higher cancer incidence rates. In other words, tweeting about cancer was most associated with a state's population and Twitter user base rather than cancer-related factors. These findings highlight an opportunity for cancer-related social media resources to be better targeted to locations that might need to be better engaged on social media.

Through hand coding of a random sample of tweets, we found that Twitter has value to stakeholders. Specifically, family members, friends of cancer patients, journalists and cancer patients/survivors were found to be the most represented user types in our data set. The latter as well as family members were most likely to tweet about personal stories, anecdotes and jokes. Though most cancer-related tweets are news-related, the amount of personal tweets is significant and establishes that Twitter is an important medium to the cancer community.

The discussion of cancer on Twitter appears to be linked to populations of educated urban professionals. Interestingly, however, social media technologies have high rates of uptake amongst lower-income, urban racial minorities — populations with long-standing health inequalities and many with higher than average cancer incidence. Therefore, there is a real opportunity for these populations who have traditional issues of access to cancer information and support to become actively engaged with social media. For example, rather than high-quality cancer centres using their social media resources to exclusively target geographically proximate populations, these institutions could

also target more distant locations that might be more in need of their messaging. As social media can remove issues of geographical access, campaigns of this sort could have a real value to locations with higher cancer incidence rates or historical barriers to cancer-related information.

References

1. Duggan M, Ellison NB, Lampe C, et al. *Social media update 2014. Pew Internet and American Life Project*. Pew Research Center, 2015, http://www.pewinternet.org/2015/01/09/social-media-update-2014/ (accessed 28 June 2016).
2. Lenhart A, Purcell K, Smith A, et al. Social media & mobile internet use among teens and young adults. *Pew Internet and American Life Project*. Pew Research Center, 2010, http://www.pewinternet.org/reports/2010/social-media-and-young-adults.aspx (accessed 28 June 2016).
3. Perkins EA and LaMartin KM. The Internet as social support for older carers of adults with intellectual disabilities. *Policy Pract Intellect Disabil* 2012; 9: 53–62.
4. Keller EG. Forget funeral selfies. What are the ethics of tweeting a terminal illness? *The Guardian*, 8 January 2014.
5. Murthy D, Gross A and Longwell S. Twitter and e-health: A case study of visualizing cancer networks on Twitter. In: *Proceedings of Information Society (i-Society), 2011 International Conference*. London, UK, 2011.
6. Murthy D. *Twitter: Social communication in the Twitter age*. Cambridge: Polity, 2013.
7. Chung DS and Kim S. Blogging activity among cancer patients and their companions: Uses, gratifications, and predictors of outcomes. *J Am Soc Inf Sci Technol* 2008; 59: 297–306.
8. Lange PG. Publicly private and privately public: Social networking on YouTube. *J Comput Mediat Commun* 2007; 13: 361–380.
9. Payne R. Frictionless sharing and digital promiscuity. *Communication and Critical/Cultural Studies* 2014; 11: 85–102.
10. Blua A. From losing weight to spotting cancer, there's a smart bra for that. Radio Free Europe/Radio Liberty. Prague, Czech Republic, 2014, http://www.rferl.org/content/feature/25218974.html (accessed 10 July 2014).
11. Memorial Hermann. Brain surgery live on Twitter. Storify, 2012, http://storify.com/memorialhermann/brain-surgery-live-on-twitter (accessed 9 January 2014).
12. Forgie SE, Duff JP and Ross S. Twelve tips for using Twitter as a learning tool in medical education. *Med Teach* 2013; 35: 8–14.
13. McNab C. What social media offers to health professionals and citizens. *Bull World Health Organ* 2009; 87: 566.
14. Keller B. Tracing digital thyroid culture: Poster. *Commun Des Q Rev* 2013; 1(4): 61.
15. Hawn C. Take two aspirin and tweet me in the morning: How Twitter, Facebook, and other social media are reshaping health care. *Health Aff* 2009; 28: 361–368.
16. Vance K, Howe W and Dellavalle RP. Social Internet sites as a source of public health information. *Dermatol Clin* 2009; 27: 133–136.
17. Thackeray R, Burton S, Giraud-Carrier C, et al. Using Twitter for breast cancer prevention: An analysis of breast cancer awareness month. *BMC Cancer* 2013; 13: 508.
18. Chaudhry A, Glodé LM, Gillman M, et al. Trends in Twitter use by physicians at the American Society of Clinical Oncology Annual Meeting, 2010 and 2011. *J Oncol Pract* 2012; 8: 173–178.
19. Victorian B. Nephrologists using social media connect with far-flung colleagues, health care consumers. *Nephrology Times* 2010; 3(1): 16–18.
20. Thompson MA, Younes A and Miller RS. Using social media in oncology for education and patient engagement. *Oncology* 2012; 26(9): 782–791.
21. Butcher L. How Twitter is transforming the cancer care community. *Oncology Times* 2009; 31(21): 36–39.
22. Hilton S, Emslie C, Hunt K, et al. Disclosing a cancer diagnosis to friends and family: A gendered analysis of young men's and women's experiences. *Qual Health Res* 2009; 19: 744–754.
23. Chou W-yS, Hunt Y, Beckjord EB, et al. Social media use in the United States: Implications for health communication. *J Med Internet Res* 2009; 11(4): 1–12.
24. Tsuya A, Sugawara Y, Tanaka A, et al. Do cancer patients tweet? Examining the twitter use of cancer patients in Japan. *J Med Internet Res* 2014; 16: e137.
25. Himelboim I and Han JY. Cancer talk on twitter: Community structure and information sources in breast and prostate cancer social networks. *J Health Commun* 2014; 19: 210–225.
26. Culotta A. Lightweight methods to estimate influenza rates and alcohol sales volume from Twitter messages. *Lang Resour Eval* 2013; 47: 217–238.

27. Dizon DS, Graham D, Thompson MA, et al. Practical guidance: The use of social media in oncology practice. *J Oncol Pract* 2012; 8: e114–e124.

28. Butcher L. Oncologists using Twitter to advance cancer knowledge. *Oncology Times* 2010; 32(1): 8–10.

29. Graham M, Hale SA and Gaffney D. Where in the world are you? Geolocation and language identification in Twitter. *Prof Geogr* 2014; 66: 568–578.

30. Burton SH, Tanner KW, Giraud-Carrier CG, et al. 'Right time, right place' health communication on Twitter: Value and accuracy of location information. *J Med Internet Res* 2012; 14: e156.

31. Hughes E, Kilmer G, Li Y, et al. Surveillance for certain health behaviors among states and selected local areas: United States, 2008. *MMWR Surveill Summ* 2010; 59(10): 1–221.

32. United States Census Bureau. *Statistical Abstract of the United States: 2012*. Washington DC: United States Census Bureau, 2012.

33. Hubspot. Top Twitter cities. Hubspot, 2009, http://tweet.grader.com/top/cities (accessed 8 October 2011).

34. Arbyn M, Castellsagué X, de Sanjosé S, et al. Worldwide burden of cervical cancer in 2008. *Annals of oncology* 2011; 22(12): 2675–2686.

35. Hanson CL, Cannon B, Burton S, et al. An exploration of social circles and prescription drug abuse through Twitter. *J Med Internet Res* 2013; 15: e189.

36. Paul MJ and Dredze M. You are what you Tweet: Analyzing Twitter for public health. *Proceedings of the fifth international AAAI conference on weblogs and cocial media*. Barcelona, Spain: The AAAI Press, California, 2011.

37. De Choudhury M, Counts S and Horvitz E. Social media as a measurement tool of depression in populations. *Proceedings of the 5th Annual ACM Web Science Conference*. Paris, France: ACM, 2013.

38. DCI Group Digital. Population of Twitter users by state. 2010, http://www.dcigroupdigital.com/digital-america/?id=999 (accessed 8 October 2011).

39. Glaser BG and Strauss AL. *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: Transaction Publishers, 2009.

40. Lyles CR, López A, Pasick R, et al. '5 mins of uncomfyness is better than dealing with cancer 4 a lifetime': An exploratory qualitative analysis of cervical and breast cancer screening dialogue on Twitter. *J Cancer Educ* 2013; 28: 127–133.

41. Rowell JL, McCarthy DO and Alvarez CE. Dog models of naturally occurring cancer. *Trends Mol Med* 2011; 17: 380–368.

42. Chew C and Eysenbach G. Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PLoS One* 2010; 5: e14118.

43. Achrekar H, Gandhe A, Lazarus R, et al. Predicting flu trends using Twitter data. *Published in the IEEE conference on Computer Communications Workshops (INFOCOM WKSHPS)*. Shanghai, PR China: IEEE, 2011, pp. 702–707.

44. Culotta A. Towards detecting influenza epidemics by analyzing Twitter messages. *Published in the SOMA'10 proceedings of the first workshop on social media analytics*. Washington DC: ACM, 2010.

45. Mislove A, Lehmann S, Ahn Y-Y, et al. Understanding the demographics of Twitter users. *Published in the fifth international AAAI conference on weblogs and social media (ICWSM-11)*. Barcelona, Spain, 2011.

46. Centers For Disease Control And Prevention. *Interactive Cancer Atlas (InCA)*. US cancer statistics: An interactive map. 2010, http://apps.nccd.cdc.gov/DCPC_INCA/DCPC_INCA.aspx (accessed 8 October 2011).

47. Watanabe M and Suzumura T. How social network is evolving?: A preliminary study on billion-scale twitter network. *Proceedings of the 22nd international conference on World Wide Web companion*. Rio de Janeiro, Brazil: International World Wide Web Conferences Steering Committee, ACM, 2013.

48. Hargittai E and Litt E. The tweet smell of celebrity success: Explaining variation in Twitter adoption among a diverse group of young adults. *New Media Soc* 2011; 13: 824–842.

49. Digital surgeons. Facebook vs. Twitter: A breakdown of 2010 social demographics. 2010, http://www.digitalsurgeons.com/Twitter-vs-twitter-infographic/ (accessed 31 January 2012).

50. Murthy D, Gross A and Oliveira D. Understanding cancer-based networks in Twitter using social network analysis. *Published in the proceedings of the 2012 IEEE sixth international conference on Semantic Computing*. California, USA: IEEE, 2011.